

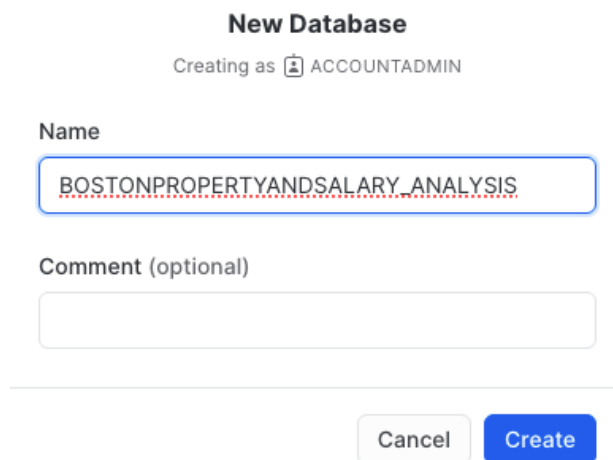
I. Background

Database and schemas are used to organize data in Snowflake. Each database contains one or more schemas. Schemas are logical groupings of database objects such as tables, views. There is no limit on the number of databases, schemas and objects can create. [1]


II. Snowflake: Load Data

In order to load the Boston Property csv, it is necessary to create a database, schema and table.

I created a database named `BostonPropertyAndSalary_Analysis`, as seen in Figure 1, with the intention of analyzing trends in Boston properties, and trends between Boston properties and salaries. After I created the `BostonPropertyAndSalary_Analysis` database, I created the schema: `Data`, to store data tables, as seen in Figure 3. I created a table named `Boston_Property_Details_2` with SQL, as seen in Figure 5, and loaded the `Boston_Property_Details_23` csv into `Boston_Property_Details_2`.



New Database

Creating as  ACCOUNTADMIN

Name

`BOSTONPROPERTYANDSALARY_ANALYSIS`

Comment (optional)

Figure 1. Create New Database: `BostonPropertyAndSalary_Analysis`

Databases

+ Database

7 Databases

Search

Source All

NAME ↑	SOURCE	OWNER	CREATED	
<div></div> BOSTON_PROPERTYANDSALARY...	Local	<div></div> ACCOUNTADMIN	5 days ago	...
<div></div> CHAIRLIFT_CONSUMER_DATA	Local	<div></div> CHAIRLIFT_ADMIN	5 days ago	...
<div></div> CHAIRLIFT_PKG	Share	<div></div> CHAIRLIFT_PROVIDER	5 days ago	...
<div></div> CHAIRLIFT_PROVIDER_DATA	Local	<div></div> CHAIRLIFT_PROVIDER	5 days ago	...
<div></div> GLOBAL_WEATHER_CLIMATE_DA...	Share	<div></div> ACCOUNTADMIN	6 days ago	...
<div></div> SNOWFLAKE	Share	—	6 days ago	...
<div></div> SNOWFLAKE_SAMPLE_DATA	Share	<div></div> ACCOUNTADMIN	6 days ago	<div></div> ...

Figure 2. Databases

New Schema

Creating as ACCOUNTADMIN

Name

DATA

Comment (optional)

☐ Managed access

Cancel

Create

Figure 3. Create New Schema: Data

BOSTON_PROPERTYANDSALARY_ANALYSIS / DATA

Schema

ACCOUNTADMIN

5 days ago

Schema Details

Tables

Privileges

ACCOUNTADMIN (Current Role)

OWNERSHIP

Group by Role

...

Create

Table

View

Stage

Storage Integration

File Format

Sequence

Pipe

Stream

Task

Function

Procedure

Dynamic Table

Figure 4. Create Table

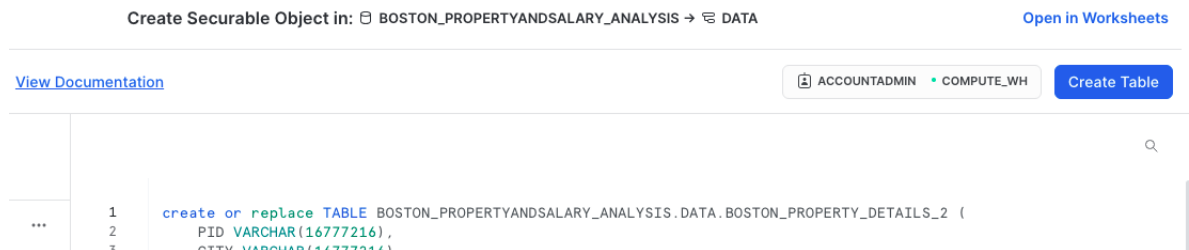


Figure 5. SQL Code to Create Table

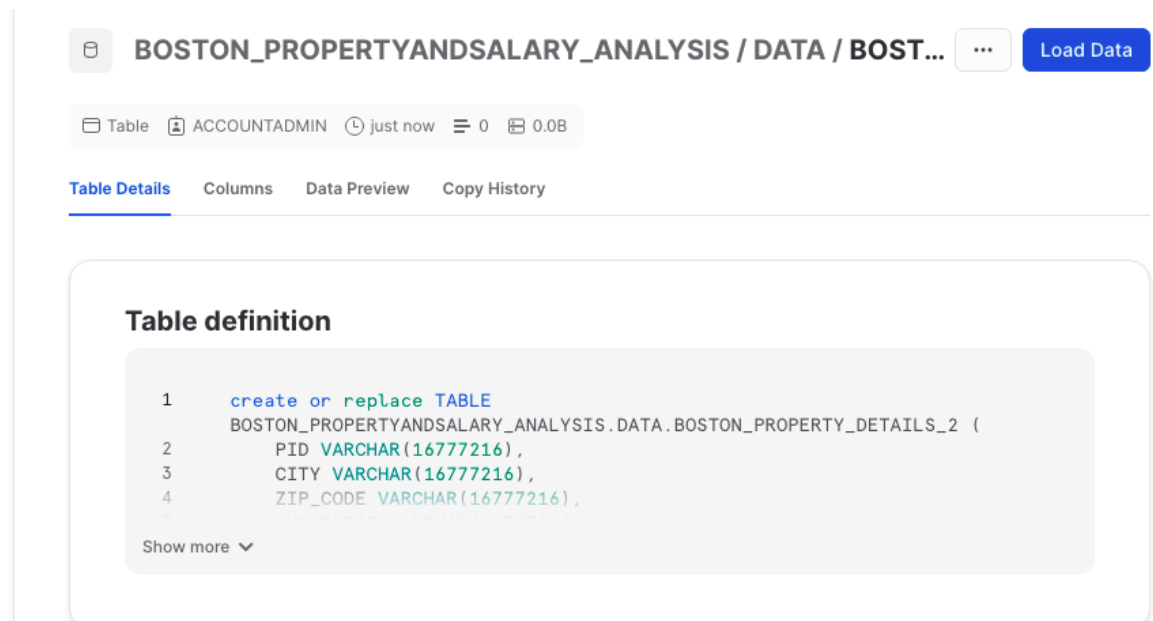


Figure 6. Boston_Property_Details_2 Table

Load Data into Table

BOSTON_PROPERTY_DETAILS_2

COMPUTE_WH

Drag and drop to upload files

or

Browse

or

Add from Stage

File size limit: 50MB

Cancel

Back

Next

Load Data into Table

BOSTON_PROPERTY_DETAILS_2

COMPUTE_WH

Boston_Property_Details_23.csv - 27.4MB

Browse

Cancel

Back

Next

Figure 7. Load Data (csv) into Boston_Property_Details Table

III. Snowflake: Clean, Transform and Analyze Data

I cleaned, transformed and analyzed the data through creating a SQL worksheet, called Clean & Transform Boston Property Data as seen in Figure 8. I have included an example of a run of the Clean & Transform Boston Property Data worksheet in Figure 9.

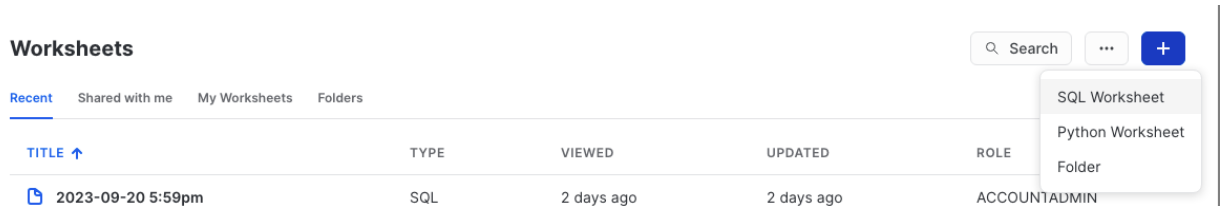


Figure 8. Create Worksheet: Clean & Transform Boston Property Data

The screenshot shows the Snowflake SQL Worksheet interface. At the top, there's a header bar with the worksheet name 'Clean & Transform Boston ...', a '+ Add' button, and a toolbar with 'ACCOUNTADMIN', 'COMPUTE_WH', 'Share', and a play button. Below the header, there's a code editor with the following SQL code:

```
1  -----
2  -- Boston Property Analysis --
3  -----
4
5  -- View data
6  SELECT *
7  FROM boston_property_details;
8
9  -- *****
10 -- DATA CLEANING -----
11 -- *****
12
13 -- Remove duplicate rows
14 DROP TABLE IF EXISTS boston_property_details_cleaned;
15 CREATE TABLE boston_property_details_cleaned AS
16 SELECT DISTINCT *
17 FROM boston_property_details;
18
19 -- View data
```

Below the code editor, there's a 'Results' tab and a 'Chart' tab. The 'Results' tab is active, showing a table with the following data:

	AVG_LAND_COST_BOSTON	AVG_BUILDING_COST_BOSTON	AVG_PRPTY_COST_BOSTON
1	376,585	1,120,621	1,500,193

On the right side of the 'Results' tab, there's a 'Query Details' panel with the following information:

- Query duration: 122ms
- Rows: 1
- Query ID: 01af401f-0404-c042-0...

At the bottom of the 'Query Details' panel, there's a table with the following data:

AVG_LAND_COST_BOSTON	#

Figure 9. Clean & Transform Boston Property Data Worksheet

IV. Appendix:

1. Snowflake: Databases, Tables & Views.
<https://docs.snowflake.com/en/user-guide/databases>